# Forecasting of The Crime Rate Using Automatic Clustering and Fuzzy Logic Relationship Method In North Sumatra

Elsyah Suhadiyah<sup>1</sup>, Sajaratud Dur<sup>2</sup>, Hendra Cipta<sup>3\*</sup>

<sup>1,2,3</sup> Department of Mathmatics, Faculty of Science and Technology Universitas Islam Negeri Sumatera Utara Medan, Indonesia

> \*Corresponding author: Email: <u>hendracipta@uinsu.ac.id</u>

#### Abstract.

Currently the crime rate is very alarming and reported in various mass and electronic media. The high crime rate in the Province of Nort Sumatra is very unsettling for the community. The purpose of this research is to get the result of forecasting the crime ratel in the 2021 - 2024 using Automatic Clustering And Fuzzy Logic Relationship (ACFLR) method. The advantage of this method is that the method has a high level of accuracy because the Mean Absolute Percentage Error (MAPE) value is relative small and the results of forecasting analysis obtained in 2021 there are 31522 cases, in 2022 are 31533 cases, in 2023 are 31574 cases and the last one in 2024 was 31602 cases. In addition, the prediction error rate MAPE obtained is 0,35 %

Keywords: Crime rate, Automatic Clustering And Fuzzy Logic Relationship.

#### **1. INTRODUCTION**

Forecasting serves to predict future event based on scientifically analyzed past data [2], [8]. Forecasting can be qualitative or quantitative. Qualitative forecasting is not in the form of numbers while forecasting is quantitative in the form of numbers and mostly show in digital form [6]-[8].

Fuzzy time series is new concept for forecasting using fuzzy logic, namely time series forecasting problems that is able to provide an explanation of fuzzy data and is presented in linguistic values and the fuzzy time series also captures patterns form past data and then uses them to forecasting future data [8]-[12]. In 1993, Song and Chissom introduced the fuzzy time series method for the first time, which was alleg-edly able to fill the short comings of the time series method [4], [9]. In 2009 Chen, Wang and pan introduced a new method, the method Automatic Clustering And Fuzzy Logic Relationship (ACFLR) [1]. Form several previous studies, it was concluded that the comparison of the ACFLR and ARIMA methods by Khairunnisa and Sunendiari showed that the lowest MAPE level was the ACFLR method and the highest was the ARIMA method. Likewise, Whib research that applies the ACFLR method for forecasting staple goods and Sihotang research in predicting gold prices and Abdy re-

search which predicts the population of Makassar, they conclude that the error rate obtained is very small or more minimum [2].

The current crime rate is very concerning and is reported in various massa and electronic media. The varying levels of crime in North Sumatra Province are very disturbing for the surrounding community. Based on information from the Central Statistics Agency, Nort Sumatra is one of the areas with the highest crime percentage in Indonesia, with a crime rate at the provincial/polda level in 2017. The North Sumatra police recorded the highest crime rate with 38867 cases and was in first pisition at the National level, then in 2018 the North Sumatra police with 32922 cases was in sec-ond position at the National level. Meanwhile, in 2019 the North Sumatra police recorded a crime rate of 30831 cases and was ranked second at the National level and in 2020 recorded 32990 cases and was ranked first Nationally [7]. In a addition, the rate of crime or violation reported by type of crime or violation in North Sumatra in 2020 was 31258 cases. Therefore, researchers will apply the Automatic Clustering And Fuzzy Logic Relationship (ACFLR) method to predict the crime rate in North Sumatra Province with the hope that the forecast results have high accuracy, so thvt they can control the crime rate in the future.

# **II. METHODS**

#### Automatic Clustering Algorithm (ACA)

Automatic Clustering Algorithm (ACA) first introduced by Chen, Wang and Pan which is a modification of the clustering algorithm. The steps can be explained as follows [1], [10], [11]:

1. Sort data from the smallest to largest and from the order of data the average difference is calculated. The formula is showed:

$$Average\_diff = \frac{\sum_{i=1}^{n-1} (d_{i+1} - d_i)}{n-1}$$
(1)

2. Converting data into cluster

Take the first data (smallest data in ascending data sequence) to the current cluster. According to the average\_diff value, determine whether the numbers in the ascending data sequence are included in the current cluster or placed in the new cluster according to the following principles:

- a. Suppose the current cluster is the first cluster, and there is only one data  $d_1$ , and let  $d_2$  the data that is adjacent to the data  $d_1$  as follows  $\{d_1\}, d_2, d_3, \ldots, d_n, d_2 d_1 \le average\_diff$  then put  $d_2$  into the main cluster that is owned by  $d_1$ . Let the new cluster for  $d_2$ , and let the new cluster that is formed where  $d_2$  form the current cluster.
- b. Assume that the current cluster is not the first cluster, and there is only one data in this current cluster. Assume that  $d_k$  is adjacent data after adjacent data, and assume that  $d_i$  is data with the largest value among existing clusters in the data before the current cluster as shown

 $\{d_1, d_2\}, \{d_3, d_4\}, \dots, \{d_i, d_j\}, \{d_k, d_l\}, \dots, \{d_r\}, \{d_s, d_l\}, \dots, \{d_{n-1}, d_n\}.$  If

 $d_k - d_j \le average\_diff$  and  $d_k - d_j \le d_j - d_i$ . Then put  $d_k$  into the current cluster in it. Otherwise form a new cluster for  $d_k$  and let the new cluster formed where  $d_k$  is included as the current cluster.

c. Assume that a current cluster is not the first cluster and there are multiple data sets in the current cluster. Assuming  $d_i$  is the largest data cluster at this time, and assuming that  $d_j$  is the newest data set after  $d_i$ , it looks like

$$\{d_1,\ldots\},\ldots,\{\ldots\},\{\ldots,d_i\},d_j,\ldots,d_n. \quad \text{If} \quad \begin{cases} d_j - d_i \leq average\_diff \\ d_q - average\_diff \\ d_q + average\_diff \end{cases} \text{ and}$$

 $d_j - d_i \leq cluster\_diff$  in the current cluster, then place  $d_j$  in the current cluster where  $d_i$  is in it. Instead, form a new cluster for  $d_j$  and leave the new cluster formed where  $d_j$  belongs to the current cluster, where cluster\_diff represents the difference in the mean distance between each pair of adjacent data in the cluster. It is assumed that the data in the current cluster is  $\{c_1, c_2, c_3, \dots, c_i, \dots, c_n\}$  and the value of cluster\_diff is formulated as follows:

$$cluster\_diff = \frac{\sum_{i=1}^{n-1} (c_{i+1} - c_i)}{n-1}$$
(2)

3. Improve the contents of the cluster

According to the cluster results obtained in step 2, adjust the contents of the cluster according to the following principles.

- a. If the cluster has more than two data, it will use the smallest data, the largest data, and then delete the remaining data.
- b. If the cluster only has two data, then it is ignored (no change)
- c. If the cluster has only one  $d_q$  data, put  $d_q$ -average\_diff and  $d_q$ +average\_diff into the cluster, then delete the  $d_q$  value from this cluster. In addition, if this happens, the cluster will need to be readjusted.

# 4. Converting clusters to intervals

Suppose the cluster results obtained from step 3 are as follows

 $\{d_1, d_2\}, \{d_3, d_4\}, \dots, \{d_i, d_j\}, \{d_k, d_l\}, \dots, \{d_r\}, \{d_s, d_t\}, \dots, \{d_{n-1}, d_n\}$ . Convert the cluster to a continuous interval by performing the following steps:

- 1) Convert the first cluster  $\{d_1, d_2\}$  into interval  $[d_1, d_2)$ .
- 2) If the current interval is  $[d_i, d_j)$  and if the current cluster is  $\{d_k, d_l\}$ , then:
  - a. If  $d_j \ge d_k$ , then convert current cluster  $\{d_k, d_l\}$  into interval  $[d_k, d_l)$ . Let  $[d_k, d_l)$  be the current interval and then let the next cluster  $\{d_m, d_n\}$  be the current cluster.
  - b. If  $d_j < d_k$ , then replace  $\{d_k, d_l\}$  into the interval  $[d_k, d_l)$  and create a new interval  $[d_j, d_k)$  between  $[d_l, d_j)$  and  $[d_k, d_l)$ . Let  $[d_k, d_l)$  be the current interval and let the next cluster  $\{d_m, d_n\}$  be the current cluster.

If the current interval is  $[d_i, d_j)$  and the current cluster is  $\{d_k\}$ , then change the current interval  $[d_i, d_j)$  to  $[d_i, d_k)$ . Let  $[d_i, d_k)$  be the current interval and let the next cluster be the current cluster.

- 3) Check the current interval and clusters repeatedly until all clusters are converted to interval form.
- 5. Partitioning interval

For each interval obtained from step 4, then divide each interval into p subinterval where  $p \ge 1$ .

#### Automatic Clustering And Fuzzy Logic Relationship (ACFLR)

The next step is to calculate the forecast value using the Automatic Clustering And Fuzzy Logic Relationship (ACFLR) method after obtained the interval using the automatic clustering algorithm. The steps are as follows [5], [14]:

- 1. The set of universes is  $U = [D_{\min}, D_{\max}]$  determined according to the existing data. Automatic Clustering Algorithm is used to create intervals from existing data and calculate the midpoint of each interval.
- 2. Fuzzification process, it is assumed that there are n intervals  $u_1, u_2$  and  $u_n$  then defines each fuzzy set  $A_i$  which  $1 \le i \le n$ , is as follows:

$$A_{1} = \frac{1}{u_{1}} + \frac{0.5}{u_{2}} + \frac{0}{u_{3}} + \frac{0}{u_{4}} + \dots + \frac{0}{u_{n-1}} + \frac{0}{u_{n}}$$

$$A_{2} = \frac{0.5}{u_{1}} + \frac{1}{u_{2}} + \frac{0.5}{u_{3}} + \frac{0}{u_{4}} + \dots + \frac{0}{u_{n-1}} + \frac{0}{u_{n}}$$

$$\vdots$$

$$A_{n} = \frac{0}{u_{1}} + \frac{0}{u_{2}} + \frac{0}{u_{3}} + \frac{0}{u_{4}} + \dots + \frac{0.5}{u_{n-1}} + \frac{1}{u_{n}}$$
(3)

- 3. Fuzzification of each data into a fuzzy set, if an interval  $u_1$  is included in the data when  $1 \le i \le n$ , then the data is fuzzified into a fuzzy set  $A_i$ . Where the fuzzy set is  $A_1, A_2, A_3, \ldots, A_n$  a fuzzy set whose variables are determined from the state of the universe.
- 4. The basis for making a fuzzy logical relationship is by fuzzification of the existing data in the previous step with the results obtained from fuzzification with years t and t + 1, each fund, then A<sub>i</sub> and A<sub>k</sub> fuzzy logical relationship is A<sub>i</sub> → A<sub>k</sub> formed where A<sub>i</sub> and A<sub>k</sub> each fund is called the current state and next states of fuzzy logical relationship.
- 5. The Defuzzification process is calculated based on the following principles:
  - a. If the fuzzification value of year t is  $A_j$  and there is only one fuzzy logical relationship in the fuzzy logical relationship group that has a current state of  $A_j$  as shown  $A_j \rightarrow A_k$  forecasting value is year t + 1 is  $m_k$ , where  $m_k$  is the midpoint of the interval  $u_k$ , and the maximum membership value of a fuzzy set  $A_k$  appears on the interval  $u_k$ .

b. If the fuzzification value in year t is  $A_j$ , and the following fuzzy logical relationship exists in the fuzzy logical relationship group having a current state as shown:

$$A_i \to Ak_1(x_1), Ak_2(x_2), \dots Ak_p(x_p) \tag{4}$$

Then the forecast value from year t + 1 is calculated as follows:

$$\frac{x_1 m_{k1} + x_2 m_{k2} \dots + x_p m_{kp}}{x_1 + x_2 \dots + x_p}$$
(5)

where  $x_i$  is derived from the fuzzy logic relation. The  $A_j \rightarrow A_{ki}$  set of fuzzy logical relations,  $1 \le i \le p$ ;  $m_{k1}, m_{k2}$  and  $m_{kp}$  is the midpoint of the interval between  $u_{k1}, u_{k2}, \ldots$  and  $u_{kp}$  and the maximum membership value of the fuzzy sets  $A_{k1}, A_{k2}, \ldots$  and  $A_{kp}$  are  $u_{k1}, u_{k2}$  and  $u_{kp}$  respectively.

## **III. RESULT AND DISCUSSION**

#### **Sample Data**

Orininal data is shown in table 1.

Table 1. Data on the Number of Reported Crimes/Violations by Type ofCrime/Violation in North Sumatra Province 2010 - 2020

Years	<b>Crime Rate</b>
2010	39.220
2011	44.104
2012	41.734
2013	43.652
2014	38.669
2015	42.324
2016	42.974
2017	22.982
2018	37.374
2019	33.653
2020	31.258

# Forecasting the crime rate in North Sumatra Province using ACFLR method

If the automatic clustering algorithm has applied, then the following interval is formed:

$u_1 = [22982, 25094.2)$	$u_5 = [35765.2, 37374)$
$u_2 = [25094.2, 29145.8)$	$u_6 = [37374, 39220)$
$u_3 = [29145.8, 33370.2)$	$u_7 = [39220, 41734)$
$u_4 = [33370.2, 35765.2)$	$u_8 = [41734, 44104)$

The next step, make intervals using the sub interval p = 8 where  $p \ge 1$  and the interval obtained is 64 intervals and calculate the mean value of each interval that has been

obtained.

Years	<b>Crime Rate</b>	Fuzzification	
2010	39.220	$A_{49}$	
2011	44.104	$A_{64}$	
2012	41.734	$A_{57}$	
2013	43.652	$A_{63}$	
2014	38.669	$A_{46}$	
2015	42.324	$A_{58}$	
2016	42.974	$A_{61}$	
2017	22.982	$A_{1}$	
2018	37.374	$A_{41}$	
2019	33.653	$A_{25}$	
2020	31.258	$A_{21}$	

Table 2. Data on the fuzzification of crime rates in North Sumatra Province

From the table above, the next step is to create a fuzzy logical relationship. Based on the fuzzy Logical relationship table, it can be seen that a fuzzy logical relationship between 2010 and 2011 can be made  $A_{49} \rightarrow A_{64}$ , by looking at the results of the fuzzification of the crime rate in 2010  $A_{49}$  and 2011  $A_{64}$  being, with  $A_{49}$  and  $A_{64}$  respectively called the current state and next state (next state) of fuzzy logical relationship.

 Table 3. Fuzzy Logical Relationship Crime Rate in North Sumatra

FLR	
$A_{49} \rightarrow A_{64}$	
$A_{64} \rightarrow A_{57}$	
$A_{57} \rightarrow A_{63}$	
$A_{63} \rightarrow A_{46}$	
$A_{46} \rightarrow A_{58}$	
$A_{58} \rightarrow A_{61}$	
$A_{61} \rightarrow A_1$	
$A_1 \rightarrow A_{41}$	
$A_{41} \rightarrow A_{25}$	
$A_{25} \rightarrow A_{21}$	
$A_{21} \rightarrow \#$	

https://ijsenet.com

The next step is to create a fuzzy logic relationship group by looking at the current state of the previous fuzzy logic relationship.

Group	FLR
Group 1	$A_1 \rightarrow A_{41}$
Group 2	$A_{21} \rightarrow \#$
Group 3	$A_{25} \rightarrow A_{21}$
Group 4	$A_{41} \rightarrow A_{25}$
Group 5	$A_{46} \rightarrow A_{58}$
Group 6	$A_{49} \rightarrow A_{64}$
Group 7	$A_{57} \rightarrow A_{63}$
Group 8	$A_{58} \rightarrow A_{61}$
Group 9	$A_{61} \rightarrow A_1$
Group 10	$A_{63} \rightarrow A_{46}$
Group 11	$A_{64} \rightarrow A_{57}$

Table 4. Fuzzy Logical Relationship Group Criminal Rate in North Sumatra

Defuzzification, this process is a process for forecasting crime in 2010, based on table 2, it can be seen that the results of the fuzzification of the crime rate in 2010  $A_{49}$  are from table 3, it can also be seen that there is a fuzzy logical relationship  $A_{49} \rightarrow A_{64}$  in (group 6) in table 4, then the results of forecasting the level of criminal in 2010 which is the median value of the interval  $u_{49}$ . Because  $u_{49} = [39220, 39534.25)$  and the median value is 39377.125 hereby the result of forecasting the crime rate in 2010 is 39377.

**Table 5.** Crime Rate Forecasting Results in North Sumatra 2010 – 2020

Years	Crime Rate (Reported)	Forecasting
2010	39.220	39.377
2011	44.104	43.956
2012	41.734	41.882
2013	43.652	43.660
2014	38.669	38.643
2015	42.324	42.178
2016	42.974	43.067
2017	22.982	23.114
2018	37.374	37.490
2019	33.653	33.522
2020	31.258	31.522

By using the Automatic clustering method and fuzzy logical relationships, the forecast for the crime rate in North Sumatra in 2021 will be 31,522 cases. For the next forecasting, the steps used are the same as the previous steps, with a note that if the forecast is for 2022, the value of the forecasting results obtained in 2021 is entered into historical data and the same for the following years. Forecasting obtained from 2021, 2022, 2023, 2024 using the Automatic clustering and fuzzy logical relationships method is 31,522, 31,553, 31,574, 31,602.

## MAPE (Mean Absolute Percentage Error)

MAPE (Mean Absolute Percentage Error) is one of the error measuring tools commonly used in determining the level of accuracy of the results of a study. The smaller MAPE level in a research result means that the accuracy of a research is getting better and better [3] The way to calculate MAPE is as follows:

$$MAPE = \frac{1}{n} \sum_{t=1}^{n} \left| PE_t \right| \tag{6}$$

$$PE_{t} = \left(\frac{Y_{t} - F_{t}}{Y_{t}}\right) \times 100 \tag{7}$$

Years	Calculation Error	Result
2010	( (39220-39377.125)/39220 )*100	0.40%
2011	( (44104-43955.875)/44104 )*100	0,34%
2012	( (41734-41882.125)/41734 )*100	0,35%
2013	( (43652-43659.625)/43652 )*100	0,02%
2014	( (38669-38643.125)/38669 )*100	0,07%
2015	( (42324 - 42178.375)/42324 )*100	0,34%
2016	( (42974 - 43067.125)/42974 )*100	0,22%
2017	( (22982-23114.012)/22982 )*100	0,57%
2018	( (37374-37489.375)/37374 )*100	0,31%
2019	( (33653-33534.887)/33653 )*100	0,39%
2020	( (31522-31522.025)/31522 )*100	0,84%
	MAPE	0,35%

Table 6. Error calculation using MAPE

# **IV. CONCLUSION**

The forecasting results that have been obtained in the analysis of forecasting the crime rate in North Sumatra Province using the Automatic clustering and Fuzzy Logic

Relationship method with a sub-interval p = 8 with a MAPE value of . for forecasting the crime rate in North Sumatra from 2021 to 2024. The results obtained in 2021 amounted to cases, in 2022 of cases, in 2023 of cases and the last in 2024 of cases.

#### REFERENCES

- A. M. Syam, Haryanensi. Metode Automatic Clustering And Fuzzy Logic Relationship Pada peramalan Tingkat Penduduk Di Kota Makassar. Jurnal Mathematics, Computations, and Statistics. 1, 2018, pp. 193–205.
- [2] I. Tri, Sugiman, Hendikawati. Perbandingan Tingkat Akurasi Metode Autimatic Clustering, Average Based Dan Markov Chain Fuzzy Time Series Pada Nilai Tukar (KURS) Rupiah. Jurnal Mathematics. 7, 2018, pp. 67 – 82.
- [3] K. Hilda, S. Siti. Perbandingan Metode Automatic Clustering And Fuzzy Logic Relationship Dan Arima Pada Peramalan Tingkat Pendaftaran Dan Tingkat Mahasiswa Baru Yang Melakukan Registrasi Di Universitas Islam Bandung. *Jurnal Statistika*. 5, 2019, pp. 40–46.
- [4] N. Ida, I. Ranggadara. Mean Absolute Percentage Error Un-tuk Evaluasi Hasil Prediksi Komoditas Laut. Jurnal of Information System. 5, 2020. pp. 250–255.
- [5] S. Sunarsan, S. Marubah. Analisis Peramalan Harga Emas Dengan Metode Automatic Clustering And Fuzzy Logic Relationship. *Jurnal Information System Development.* 3, 2018, pp. 104–115.
- [6] S. Sandu. Dasar Metodologi Penelitian. Yogyakarta: Literasi Media Publishing, 2018.
- [7] V. Muhammad, Muhammad. Implementasi Metode Fuzzy Time Series Untuk Peramalan Jumlah Pengunjung Di Benteng Fort Rotterdam. Jurnal of Statistic and Its Application on Teaching and Research. 2, 2020, pp. 1-12.
- [8] B. Sivanagaleela and S. Rajesh, Crime Analysis and Prediction Using Fuzzy C-Means Algorithm, 3<sup>rd</sup> International Conference on Trends in Electronics and Informatics (ICOEI), 2019, pp. 595-599, <u>https://doi.org/10.1109/ICOEI.2019.8862691</u>
- [9] R. Zhang, B, Ashuri, & Y, Deng. A novel method for forecasting time series based on fuzzy logic and visibility graph. *Adv Data Anal Classif.* 11, 2017, pp. 759–783, <u>https://doi.org/10.1007/s11634-017-0300-3</u>
- [10] M. Mittal, L. M, Goyal, L.M., Monitoring the Impact of Economic Crisis on Crime in India Using Machine Learning. *Comput Econ.* 53, 2019, pp. 1467–1485. <u>https://doi.org/10.1007/s10614-018-9821-x</u>
- [11] B. Shivani, D. Mehta, Rushabh D. Doshi Automatic Clustering Crime Region Prediction Model using Statistical Method in Data Mining. *International Journal of Engineering* and Technical Research. 9, https://doi.org/10.17577/IJERTV9IS040653
- [12] M. Melgarejo, C. Rodriguez, D. Mayorga, D., & N. Obregón, N. Time Series from Clustering: An Approach to Forecast Crime Patterns. In A. Sadollah, & C. M. Travieso-Gonzalez (Eds.), Recent Trends in Artificial Neural Networks-from Training to Prediction. IntechOpen.2019. <u>https://doi.org/10.5772/intechopen.89561</u>

- [13] E. Doynikova, E. Novikova, I. Kotenko. Attacker Behaviour Forecasting Using Methods of Intelligent Data Analysis: A Comparative Review and Prospects. Information. 11, 2020, pp. 168. <u>https://doi.org/10.3390/info11030168</u>
- [14] K. Nithyakumar, S.Preetha, R. Reshma, S. Aarthi. Crime Detection and Prevention Using Fuzzy Control Logic. SSRG International Journal of Computer Science and Engineering (SSRG-IJCSE). 2020, pp. 30.
- [15] W. Weina, L. Xiaodong, Fuzzy forecasting based on automatic clustering and axiomatic fuzzy set classification, *Information Sciences*, 294. 2015, pp. 78-94. <u>https://doi.org/10.1016/j.ins.2014.09.027</u>.
- [16] Q. Wangren, Z. Ping, W. Yanhong. Fuzzy Time Series Forecasting Model Based on Automatic Clustering Techniques and Generalized Fuzzy Logical Relationship, *Mathematical Problems in Engineering*, 8, 2015. <u>https://doi.org/10.1155/2015/962597</u>